

--	--	--	--	--	--	--	--	--	--

MULTIMEDIA UNIVERSITY

FINAL EXAMINATION

TRIMESTER 1, 2022/2023 (TERM ID 2210)

BMD3024 – MULTIVARIATE DATA ANALYSIS
(All sections / Groups)

10 FEBRUARY 2023
9.00 a.m. - 11.00 a.m.
(2 Hours)

INSTRUCTIONS TO STUDENTS

1. This question paper consists of **TWELVE (12)** printed pages with **FOUR (4)** structured questions including the cover page.
2. Marks are shown at the end of each question.
3. Write all your answers in the answer booklet provided.
4. Answer **ALL** questions.

Question One (25 marks)

- (a) Assess the meanings of the values that correlation statistic r can take. (5 marks)
- (b) Illustrate the **TWO (2)** assumptions in conducting a chi-square test, that impacts on the analysis of output of the test. (8 marks)
- (c) A fashion student was interested in factors that predicted the salaries of catwalk models. She collected data from 231 models. For each model she asked them their salary per day (*salary*), their age (*age*), their length of experience as a model (*years*), and their industry status as a model as their percentile position rated by a panel of experts (*beauty*). She then used Linear Regression to see which variables predict a model's salary.

Table 1: Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.429 ^a	.184	.173	14.57213

- (i) Interpret **Table 1: Model Summary**. Why is the adjusted R^2 lower than the unadjusted R^2 ? (4 marks)

Table 2: ANOVA^a

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	10871.964	3	3623.988	17.066	<.001 ^b
	Residual	48202.790	227	212.347		
	Total	59074.754	230			

a. Dependent Variable: Salary per Day (£)

b. Predictors: (Constant), Attractiveness (%), Number of Years as a Model, Age (Years)

- (ii) Assess **Table 2: ANOVA**. Is the F -statistic significant at 5 percent level? What does this indicate? (4 marks)

Continued...

Table 3: Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95.0% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	-60.890	16.497		-3.691	<.001	-93.396	-28.384		
	Age (Years)	6.234	1.411	.942	4.418	<.001	3.454	9.015	.079	12.653
	Number of Years as a Model	-5.561	2.122	-.548	-2.621	.009	-9.743	-1.380	.082	12.157
	Attractiveness (%)	-.196	.152	-.083	-1.289	.199	-.497	.104	.867	1.153

a. Dependent Variable: Salary per Day (\$)

- (iii) According to Table 3: Coefficients, assess whether the variables are significant in the prediction at 5 percent level.

(2 marks)

- (iv) According to Table 3: Coefficients, assess which variable has worse multicollinearity scores than others.

(2 marks)

Continued...

Question Two (25 marks)

Piff, Stancato, Côté, Mendoza-Dentona, & Keltner (2012) used the behaviour of drivers to investigate the claim that people of a higher social class are more unpleasant. They classified social class by the type of car (*Vehicle*) on a five-point scale and observed whether the drivers cut in front of other cars at a busy intersection (*Vehicle_Cut*). A logistic regression was carried out using SPSS to determine whether social class predicts whether a driver cut in front of other vehicles.

Table 4: Case Processing Summary

Unweighted Cases ^a		N	Percent
Selected Cases	Included in Analysis	274	100.0
	Missing Cases	0	.0
	Total	274	100.0
Unselected Cases		0	.0
Total		274	100.0

a. If weight is in effect, see classification table for the total number of cases.

- (a) Assess **Table 4: Case Processing Summary** to determine if there are any missing data. (2 marks)

Continued...

Table 5: Dependent Variable Encoding

Original Value	Internal Value
Did Not Cut Off	0
Cut Off	1

Table 6: Classification Table^{a,b}

	Observed		Predicted		
			Did they cut vehicle off at the intersection?		Percentage Correct
			No	Yes	
Step 0	Did they cut vehicle off at the intersection?	No	240	0	100.0
		Yes	34	0	.0
	Overall Percentage				87.6

a. Constant is included in the model.

b. The cut value is .500

- (b) Interpret **Table 5: Dependent Variable Encoding** and **Table 6: Classification Table** to determine the coding of the two categories of drivers' actions, the number of drivers in each of the two categories, and the overall percentage when the **base model (Step 0)** is applied.

(4 marks)

Table 7: Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 0 Constant	-1.954	.183	113.740	1	<.001	.142

- (c) Assess **Table 7: Variables in the Equation** to determine if the base model is able to predict make predictions that are significantly different than by chance (50-50).

(2 marks)

Continued...

Table 8: Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	4.161	1	.041
	Block	4.161	1	.041
	Model	4.161	1	.041

- (d) Interpret **Table 8: Omnibus Tests of Model Coefficients** and determine if logistic regression model that includes the independent variable 'social class by the type of car' (*Vehicle*) is significant.

(4 marks)

Table 9: Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	201.334 ^a	.015	.029

a. Estimation terminated at iteration number 5 because parameter estimates changed by less than .001.

- (e) Interpret **Table 9: Model Summary**.

(4 marks)

Table 10: Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	4.078	3	.253

- (f) Based on **Table 10: Hosmer and Lemeshow Test**, assess whether the null hypothesis that predictions made by the model, fit well with observed group memberships is rejected.

(2 marks)

Continued...

Table 11: Classification Table^a

	Observed		Predicted		
			Did they cut vehicle off at the intersection?		Percentage Correct
			Did Not Cut Off	Cut Off	
Step 1	Did they cut vehicle off at the intersection?	No	240	0	100.0
		Yes	34	0	.0
	Overall Percentage				87.6

a. The cut value is .500

- (g) Interpret Table 11: Classification Table to determine the number of drivers predicted in each of the two categories, the overall percentage when the model with one predictor (Step 1) is applied.

(2 marks)

Table 12: Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I. for EXP(B)	
								Lower	Upper
Step 1 ^a	Vehicle Status (1 = Low, 5 = High)	.365	.184	3.938	1	.047	1.441	1.005	2.067
	Constant	-3.163	.662	22.835	1	<.001	.042		

a. Variable(s) entered on step 1: Vehicle Status (1 = Low, 5 = High).

- (h) Assess from Table 12: Variables in the Equation if the predictor *Vehicle Status* is a significant in determining whether the vehicle cuts off other vehicles at intersection.

(2 marks)

- (i) Based on Table 12, compute the Percentage Change in Odds of predicting correctly whether a certain vehicle will cut off others at the intersection, if *Vehicle Status* is known.

(3 marks)

Continued...

Question Three (25 Marks)

- (a) A researcher conducted a Principal Component Analysis on twenty variables. Four variables, namely, v10, v15, v18, and v20 were removed after analysing the correlation matrix. The following tables are based on analysis of the remaining sixteen variables.

Table 13: KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		.712
Bartlett's Test of Sphericity	Approx. Chi-Square	680.630
	df	120
	Sig.	<.001

- (i) Interpret **Table 13: KMO and Bartlett's Test** to determine whether factor analysis can be carried out.

(5 marks)

Table 14: Total Variance Explained

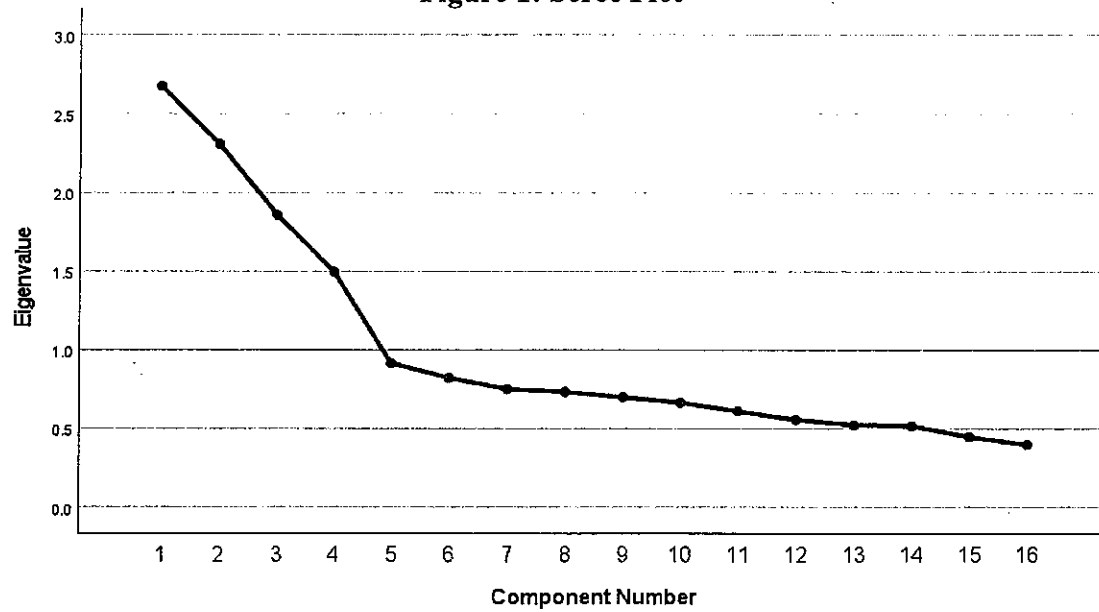
Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	2.675	16.721	16.721	2.675	16.721	16.721	2.293	14.332	14.332
2	2.308	14.427	31.148	2.308	14.427	31.148	2.208	13.798	28.130
3	1.857	11.609	42.757	1.857	11.609	42.757	2.198	13.738	41.868
4	1.496	9.353	52.110	1.496	9.353	52.110	1.639	10.242	52.110
5	.916	5.726	57.836						
.	.	.	.						
.	.	.	.						
.	.	.	.						

Extraction Method: Principal Component Analysis.

- (ii) Based on **Table 14: Total Variance Explained**, use the latent root criterion of retaining factors to interpret how many factors will be retained.

(2 marks)

Continued...

Figure 1: Scree Plot

- (iii) Based on **Figure 2: Scree Plot**, use the Scree Test criterion of retaining factors to interpret how many factors will be retained. (2 marks)
- (iv) Based on your results in (ii) and (iii), assess how many factors will be retained. (2 marks)

Continued...

- (b) In the Star Trek television series, the crew members of the spaceship wear different coloured uniforms to identify the crew member's work area – blue, red and gold. The audience watching the television series noticed that those crew members who wear red shirts have the unfortunate reputation of dying more often than those who wear gold or blue shirts, so much so that online videos and Reddit threads analysing the probability of a red-shirted crew member not surviving exists!

To confirm whether there is a correlation between red shirted crew members and certain death, the Chi-square test of independence is carried out on the number of crew members wearing blue, gold and red shirts (*Colour*), and their *Status* (Alive or Dead).

Table 15: Colour * Status Crosstabulation

			Status		Total
			Alive	Dead	
Colour	Blue	Count	129	7	136
		Expected Count	123.3	12.7	136.0
		% within Colour	94.9%	5.1%	100.0%
		% within Status	33.1%	17.5%	31.6%
		% of Total	30.0%	1.6%	31.6%
	Gold	Count	46	9	55
		Expected Count	49.9	5.1	55.0
		% within Colour	83.6%	16.4%	100.0%
		% within Status	11.8%	22.5%	12.8%
		% of Total	10.7%	2.1%	12.8%
	Red	Count	215	24	239
		Expected Count	216.8	22.2	239.0
		% within Colour	90.0%	10.0%	100.0%
		% within Status	55.1%	60.0%	55.6%
		% of Total	50.0%	5.6%	55.6%
Total		Count	390	40	430
		Expected Count	390.0	40.0	430.0
		% within Colour	90.7%	9.3%	100.0%
		% within Status	100.0%	100.0%	100.0%
		% of Total	90.7%	9.3%	100.0%

- (i) Interpret the **Table 15: Colour * Status Crosstabulation** above for the numbers and percentages of each category under the variable *Colour* according to their *Status*.

(4 marks)

Continued...

- (ii) Compute the following odds based on **Table 15**:
- A. A crew member is Dead if they are wearing Red Uniform. (2 marks)
- B. A crew member is Dead if they are wearing Blue Uniform. (2 marks)
- (iii) Compute and interpret the Odds Ratio of a Crew member status being Dead if they are wearing Red uniform versus being Dead if they are wearing Blue uniform. (2 marks)

Table 16: Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	6.189 ^a	2	.045
Likelihood Ratio	6.132	2	.047
N of Valid Cases	430		

a. 0 cells (0.0%) have expected count less than 5. The minimum expected count is 5.12.

- (iv) Based on **Table 16: Chi-Square Tests**, assess whether the assumption for chi-square is met. (2 marks)
- (v) Interpret **Table 16: Chi-Square Tests**. Is there an association between the variables *Colour* and *Status*? (2 marks)

Continued...

Question Four (25 Marks)

- (a) Sketch **ONE (1)** path diagram that shows the following hypothesized structural relationships amongst **FIVE (5)** factors: *A*, *B*, *C*, *D*, and *E*; and **TWO (2)** mediating variables, *Income* and *Gender*.

- Factors *A* to *D* are independent factors.
- *E* is a dependent factor.
- Factors *A*, *B*, and *C* are made up of **FIVE (5)** variables each.
- Factors *D*, and *E* are made up of **FOUR (4)** variables each.
- *Income* mediates the relationship between factor *A* and factor *E*.
- Factor *A* is also directly related to factor *E*.
- Factors *B*, *C*, and *D* are related to factor *E*.

(22 marks)

- (b) In the path analysis diagram drawn for question (a), illustrate all the following hypotheses and relationships between factors in the diagram:

- H₁: Factor *A* is positively related to factor *E*.
H₂: Factor *B* is positively related to factor *E*.
H₃: Factor *C* is negatively related to factor *E*.
H₄: Factor *D* is negatively related to factor *E*.
H₅: Factor *D* is positively related to *Income*
H₆: *Income* is negatively related to factor *E*.

(3 marks)

End of Page.

